Received 12 August 2022, accepted 31 August 2022, date of publication 5 September 2022, date of current version 13 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3204280

RESEARCH ARTICLE

Skin Lesion Segmentation Using Recurrent Attentional Convolutional Networks

PENG CHEN [1], SA HUANG [2], AND QING YUE [3]

[1]Department of Pediatrics, The Second Hospital of Jilin University, Changchun 130041, China
[2]Department of Radiology, The Second Hospital of Jilin University, Changchun 130041, China
[3]Department of Software Engineering, Jilin Normal University, Siping 136000, China

Corresponding author: Sa Huang (huangsa@jlu.edu.cn)

ABSTRACT Accurate segmentation of lesion region from skin lesion images can provide guidance for medical experts to make an early diagnosis of skin cancer. In this study, we construct Recurrent Attentional Convolutional Networks (O-Net), which exploits the skin lesion's attention class feature with a recurrent O-shape structure, in an iterative refinement strategy for skin lesion image segmentation. Inspired by the recently proposed attention class feature network, we integrate the attention class feature module into the proposed networks. The O-Net, with recurrent unit to iteratively refine the segmentation result, is designed to extract attention feature information and enable coarse-to-fine feature representation by iteratively integrating attention feature maps into network. Furthermore, O-Net calculates the attentional class feature by extracting attention information from the coarse segmentation result. Two currently popular datasets ISIC-2017 and PH2 are employed to explore the validity of our proposed model. The study provides detailed comparisons of our proposed network, the attention class feature network and Recurrent U-Net. O-Net achieved Dice coefficient by 87.04% on the ISIC-2017 dataset, 92.12% on the PH2 dataset with corresponding Jaccard indices of 80.36% and 86.15%, respectively, on the same dataset, which exhibits competitive performance for skin lesion image segmentation in results. The visual results also shown that more detailed tissues are extracted by O-Net than other methods.

INDEX TERMS Skin lesions, automatic image segmentation, O-Net, melanoma, attention class feature, lesion segmentation.

I. INTRODUCTION

Skin cancer is a particularly fatal and common disease all over the world. In the United States alone, more than 5 million new skin cancer patients appear every year. [1] The American Cancer Society estimates that in 2020, about 100,350 newly melanoma patients will be diagnosed in the United States, and 6,850 people (2,240 women and 4,610 men) will die from the disease. [2] The research shows the survival rate over 90% for patients who detect melanoma early [3], therefore, early discovery of skin cancer is indispensable for subsequent treatment that can prevent serious lesions.

Traditionally, experienced dermatologists by observing dermoscopy or non-dermoscopy images to diagnose the early skin cancer. [4] For dermatologists, however, diagnosing large numbers of patients is often seen as a complex and tedious process. The Computer-Aided Diagnosis (CAD) system has brought new vitality to the procedure, which was designed to raise inspection efficiency and reduce the work stress of dermatologists. At present, CAD has become an essential auxiliary facility in many screening sites and hospitals by complementation of its unique advantages with physicians to form efficient and reliable detection and diagnosis results. [5] The CAD system for skin cancer detection mainly includes the following basic aspects: target image acquisition, correlation preprocessing, lesion image segmentation, region feature extraction, and results classification. [6], [7] Lesion image segmentation is an important part of skin cancer diagnosis because it locates the exact lesion mask, which is the basis for accurate feature extraction and disease classification later. [8] However, accurately dividing the lesion region from the normal skin is a sophisticated task as a result of poor

The associate editor coordinating the review of this manuscript and approving it for publication was Fahmi Khalifa.

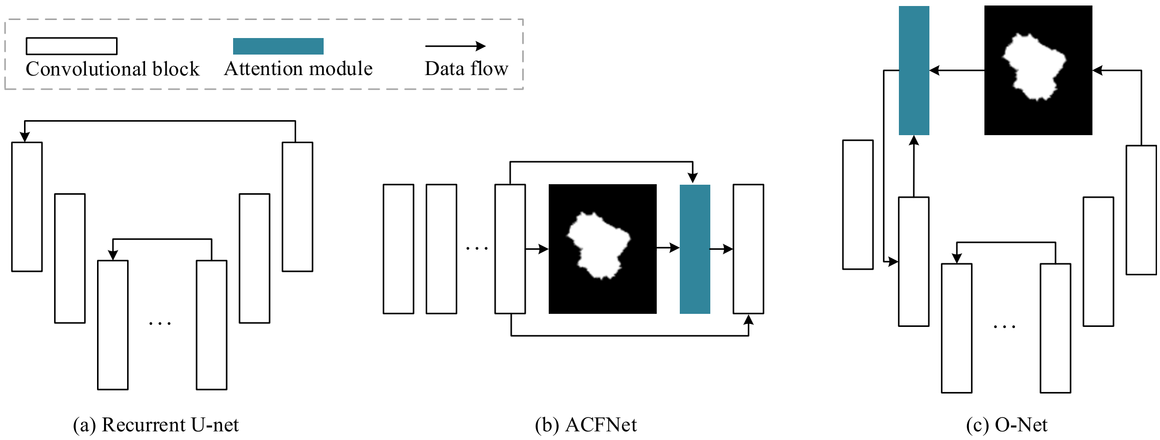

FIGURE 1. The comparison between O-Net network and the other two networks. (a) Recurrent U-net. The recurrent units achieve an iteratively updating of the internal network's states, and refine the segmentation mask simultaneously. (b) ACFNet utilize attention mechanism to obtain class-level context and generate segmentation results. (c) O-Net. Compared with the above methods, O-Net integrate the attention mechanism into the recurrent unit, so as to iteratively integrating attention maps into network and improve the segmentation results.

contrast, difference in skin color, skin aberrations, physical location of lesion and non-uniform lighting. [9] Deal with the interference, therefore, to obtain accurate lesion image segmentation results is very critical to promote CAD assistant ability.

Deep learning has shown remarkable ability in the task of processing medical images for the past few years. It is capable of exploring deep features by building network structures that include convolution, pooling and nonlinear functions, appearing impressive capabilities to excavate abstract representations from raw annotation data. Compared with traditional Convolutional Neural Network (CNN), networks formed by complementing the advantages of convolutional neural network and Recurrent Neural Network (RNN), such as U-Net [10] containing recurrent structure, have attracted more and more attention benefit from its speciality to acquire a coarse-to-fine representation. In addition, recent research has identified one of the most successful method to improve the performance is adding attention block to exploiting richer context. In the study, we constructed an attention-based mechanism network named Recurrent Attentional Convolutional Networks (O-Net) that effectively improves the capacity of the convolutional neural network to distinguish each complex voxel in the skin lesion image. It is designed as a circular closed network that coarse segmentation results are fused by the attentional module with the high-level feature map, Therefore, the context information is captured and guided throughout the segmentation network during the iteration process, so as to obtain more accurate segmentation. Furthermore, inspired by recently proposed attentional class feature network (ACFNet) [11], we improve the skin lesion image segmentation by using attention class feature module that explores class level context. ACFNet and Recurrent U-Net [12] are used for comparison. We trained all the networks from the ground up and analyzed the experimental results in detail. Figure 1 shows the simple comparison of three models.

The main contributions of the paper are as follows: (1) We first defined the concept of recurrent attentional, which represents to refine attention mechanism from the iterative process, to help network learn more accurately features and achieve a coarse-to-fine segmentation performance. (2) We proposed a recurrent attention segmentation structure with iterative refinement, named Recurrent Attentional Convolutional Networks (O-Net), equipped with an attention feature fusion module (AFFM), which containing both the class center block and the class attention block, so as to exploit class context information of coarse segmentation results to improve the final lesion segmentation result. (3) Our novel O-Net achieves accurate skin lesion segmentation results on two publicly available datasets, which is significantly competitive with other current advanced methods, especially in skin regions with different aberrations and artifacts.

In Section 2, the related work is given a brief literature review. Section 3 describes two publicly available datasets used to analyze model performance. The section also provides a briefly introduction of the proposed O-Net, ACFNet and Recurrent U-Net architecture, as well as the employed evaluation metrics. Section 4 presents experimental results, where evaluate the capability of our constructed network on two public skin lesion datasets. Discussion and conclusion are contained in Section 5 and Section 6.

II. RELATED WORKS

The purpose of skin lesion image segmentation is to extract and recognize lesion regions in skin images. With the popularity of non-invasive imaging techniques in skin cancer diagnosis, various relevant practical methods have been implemented to the skin lesion image segmentation. With reference to the strategies used, segmentation methods can be summarized as two main classes: learning-based method and non-learning-based method. Learning- based approaches learn from hand-crafted or automatically learned features to

TABLE 1. The summarization of several representative methods.

Methodology		Method	Year	Progressiveness	Limitations
Non-learning-based Methods		Cavalcanti et.al [13]	2011	Achieved Standard camera images segmentation by a sequence of steps	Multi-stages need to improved and test more extensively
		TDLS [17]	2014	Utilize joint statistical texture distinctiveness metric and a texture-based region classification algorithm to achieve segmentation	Lack of analysis on the image quality and scale
		Pathan et.al [14]	2018	Achieved hair detection considering dermoscopic hair and lesion segmentation by a chroma-based geometric deformable model	Neglect the color information of skin images
Learning-based Methods	Hand-crafted Features Based Methods	ISL-MSCA [18]	2016	Utilize multi-scale superpixel based cellular automata to achieve automated segmentation via image-wise supervised learning	The algorithm requires multiple separate stages
		Muhammad et.al [19]	2018	Multiple stages utilizing hybrid and other techniques to extract and fuse features and further used for selection and classification	Only considering some basic traits of images, such as color, texture and shape
		3-D skin lesion reconstruction [20]	2017	Considering the estimated depth of skin lesions like 3-D features, as well as the regular color, texture, and 2-D features	The depth estimation technique utilized is naïve and need to be further improved
	Feature Learning-Based Methods	MS-UNet [22]	2019	Employing multi-stage Unets to integrate low-level context information and supervised by minimizing a weighted Jaccard distance loss function	Lack of evaluations on large-scale skin datasets
		Dermonet [24]	2019	Take advantage of high-level features extracted in varying scales and resolutions	The generalization on standard segmentation benchmarks need to be improved
		MB-DCNN [26]	2020	Achieved segmentation and classification simultaneously by a coarse segmentation network and an enhanced segmentation network	Not an end-to-end framework, cannot optimize the segmentation and classification networks jointly
		SMOTE-based data augmentation [27]	2021	The proposed data augmentation technique create synthetic images to extent the dataset and solve the class imbalance problem	Lack of the analysis of this data augmentation method on the deep learning architectures
		ASCU-Net [28]	2021	The designed multi-attentive module which combined the gate, spatial and channel attention enhance the ability of the model in capturing the contextual information and the spatial correlation	Simply fusing different attention features without setting parameters in detail or fine-tuning
		Dense-shuffle attention U-Net [31]	2022	The dense attention gates and the shuffle attention module extracts high-level information and meanwhile integrate	The vital low-level features from the encoder stage are ignored
		MSAU-Net [29]	2022	Utilizing the multi-scale attention mechanism and long short term memory structure to capture hierarchical representation and discriminative features	Largely depends on the accuracy of the annotations, and the several components of the model increase the need of computational resources
		DenseNet77-based UNET [30]	2022	Improve the segmentation performance benefit from the robust feature extraction power	Not efficient to diagnose melanoma moles under intense intensity variations

obtain the fitting weight parameter and generate the prediction results from using parameter, it automatically identifies the category to which the data belongs. Following is a brief review of lesion segmentation in skin images from these two aspects. We also make a summarization of several representative methods which shown in Table1.

A. NON-LEARNING-BASED METHODS

Non-learning-based approaches that rely on spatial color distribution of the lesion image have been popular in the early stage for its widespread practicality, such as histogram thresholding, clustering, edge-based, etc. Cavalcanti *et al.* [13] proposed using rich preprocessing steps to assist skin lesion segmentation based on Otsu's thresholding method. Pathan *et al.* [14] proposed an automatic edge curve evolution approach for segmentation without the need to define the specific initial contour. Goceri and Gunay [15] used the rough segmentation and fine segmentation of two steps to implement the automatic curve evolution segmentation algorithm. They took full advantage of the color information of the lesion area in the rough segmentation process, which provided help to minimize the active contour in the fine segmentation. Zhou et al. [16] improved segmentation results by achieving a gradient vector flow (GVF) algorithm based on mean shift relied on balancing the mean difference of all the gradients and the current gradient vector. Glaister *et al.* [17] proposed a novel segmentation method to distinguish the lesion region from skin lesion image by identifying the occurrence of the constructed representation texture distribution. This method significantly improves segmentation accuracy, but it is sensitive to the complex lesion and skin texture which can lead to poor performance.

B. LEARNING-BASED METHODS

Significantly, non-learning-based approaches involve a great deal of preprocessing. They generally did not perform well on skin lesion image in different scenes, because it is usually impossible to find a suitable algorithm to deal with all intensity variability noise. To address this limitation, many learning-based methods achieve better performance on lesion images with complex intensity variations through their own advantages. For the purpose of segmentation, two processing stages are needed: firstly, the feature vectors of pixels need to be extracted; Then, the extracted vectors need to be mapped to the corresponding category label. On the basis of the approaches adopted in the different feature extraction stage, learning-based approaches are divided into hand-crafted features based approaches and feature learning based approaches in this study.

1) HAND-CRAFTED FEATURES BASED METHODS

Based on the cellular automata (CA) model, Bi *et al.* [18] constructed a practical framework for skin lesion automatic segmentation. They used the image-wise supervised learning (ISL) in the process of seed initialization to achieve an optimal probabilistic map. Nasir *et al.* [19] adopted SVM to classify combined lesion features ground on uniform distribution and active contour. Satheesha *et al.* [20] proposed a lesion segmentation algorithm based on adaptive snake technique. Each voxel expressed by intensity and spatial characteristics was classified by SVM AdaBoost, and bag-of-features (BoF) classifier.

2) FEATURE LEARNING BASED METHODS

In recent years, deep learning has gradually demonstrated its extraordinary value in many practical application scenarios. Effective feature hierarchies can be learned from the skin lesion images and labels, rather than manually capturing the features. Accordingly, Yu *et al.* [21] constructed a multilayer convolution neural network framework for the accurate skin lesion segmentation. For purpose of capture more discriminative and richer features, the proposed method combined U-Nets and context information fusion structure (CIFS) to capture multi-scale information about each pixel. Tang *et al.* [22] presented a CNN-based method which employing multi-stage U-Nets to segment the skin lesion. Salimi *et al.* [24] constructed a network namely DermoNet with structure included encoder and decoder for skin lesion segmentation. Bi *et al.* [25] extracted the features of skin lesions in the process of adversarial learning by using the generative adversarial networks (GANs) to promote the capability of segmentation. Xie *et al.* [26] constructed a MB-DCNN architecture to further improve the performance by combining the segmentation and classification of lesion regions. Olusola *et al.* [27] utilized a covariant SMOTE and proposed a data augmentation technique to generate new skin melanoma samples so as to solve the problem of class imbalance. The ASCU-Net [28] focus on the combination of the attention mechanism to capture the spatial and contextual correlation to fusing the important information of objects and achieve segmentation performance with highly reliability and robustness. MSAU-Net [29] proposed a multi-scale attention U-net to selectively adjust the hierarchical representations and utilized a long short-term memory structure to capture the discriminative features. Marriam Nawaz *et al.* [30] proposed a DenseNet77-based UNET to improve the feature extraction power of the encoder unit of UNET and enhance the ability of the model in segmenting small skin lesions. Dense-shuffle attention U-Net [31] proposed a multiple-attention-based neural network to extract high-level information and introduced a shuffle-attention unit to integrate channel and spatial attention.

In general, deep learning approach brings a new way of solving problems different from traditional non-learning-based and hand-crafted features based approaches. In this research, we constructed a novel network iteratively fusing global attention information to accomplish the task of effective automatic skin lesion images segmentation.

III. MATERIALS AND METHODS

The target of our study is to design a fully convolutional network of automatic segment skin lesion region in skin dermoscopy images. Inspired by Recurrent U-Net [12] and attentional class feature network (ACFNet) [11], we present a novel network named Recurrent Attentional Convolutional Networks (O-Net) for lesion images segmentation work.

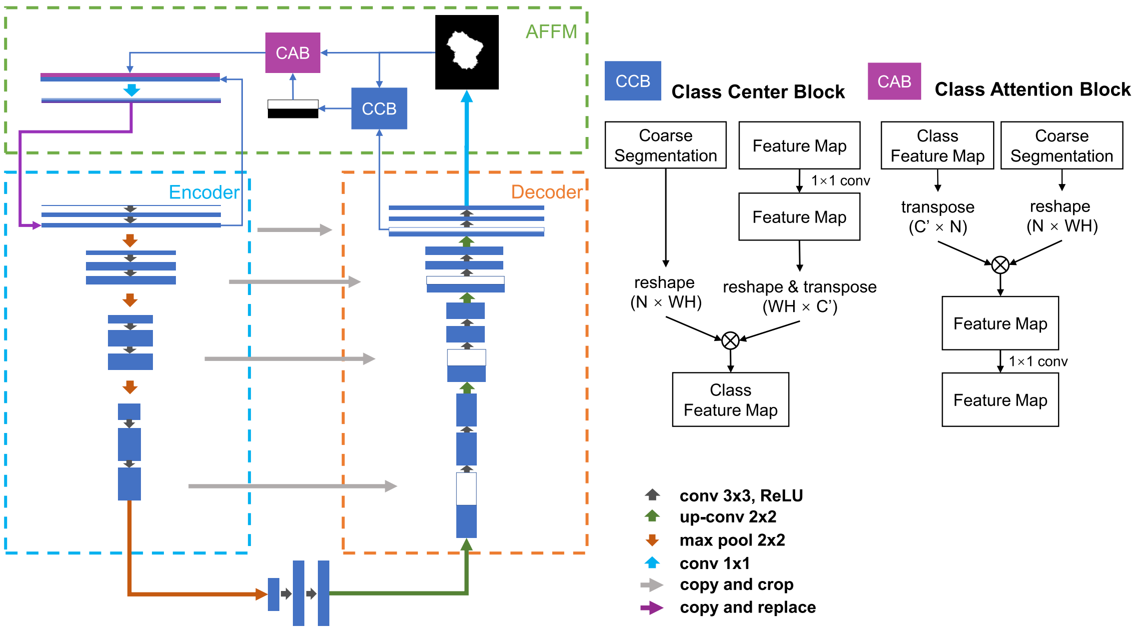

FIGURE 2. The framework of the proposed O-Net. The overall architecture consists of encoder, decoder and the attention feature fusion module (AFFM). The graph on the right gives the details of two main components in AFFM: the Class Center Block (CCB), and the Class Attention Block (CAB), which captures the context information of each pixel and obtains attention features of lesion to improve the segmentation performance.

The proposed method is designed to combine the advantages of both attentional class feature module and Recurrent U-Net architecture. In this section, we will present a detailed description of the proposed method, as well as giving a brief introduction of the two other networks.

A. DATASETS AND MATERIAL

To measure the performance of our proposed network architecture, we trained and validated on publicly available dataset namely ISIC-2017 [32] and tested on the ISIC-2017 and PH2 [33] dataset, respectively. In 2017, the ISIC-2017 dataset was aggregated and released by the International Skin Imaging Collaboration (ISIC). The number of train sets, validation sets and test sets were 2000, 100 and 600, respectively, for each, group of data contains the original dermoscopy images of skin lesion and the corresponding lesion region annotations. The original images are RGB images in JPEG format with a resolution range of 566×679 to 4499×6748 , and the lesion annotations are binary mask images in PNG format with the same resolution. The PH2 dataset was established by a collaboration of several research institutions.

B. RECURRENT ATTENTIONAL CONVOLUTIONAL NETWORKS (O-NET)

Inspired by Recurrent U-Net and attentional class feature network (ACFNet), we design a novel network named Recurrent Attentional Convolutional Networks (O-Net) for skin lesion images segmentation process. The constructed network is composed of a symmetrical encoders and decoders architecture on both sides, and the output coarse segmentation results as feature attention maps to reintegrate into the network through attention class feature module on top. The new model is trained to reintegrate global attention information, and each pixel is trained to selectively perceive different class centers of the whole scene, both of which help achieve accurate segmentation.

Figure 2 describes the proposed network architecture of O-Net. The integrated internal design of the class center block and class attention block is given by the interpretation on the right side. The architecture consists of three parts: feature encoder (left side), feature decoder (right side) and attentional feature fusion module (top side). The input skin lesion image is passed through the encoder and decoder, which are designed in U-Net, to produce coarse segmentation result Y_1 with the spatial size of $W \times H \times C$. The attentional feature fusion module guided by coarse segmentation result Y_1 extracts the attention class features between classes and integrates them into the encoder part. The feature information is iterated through encoder, decoder and feature fusion module in turn to further form the segmentation result $Y_2, Y_3 \cdots Y_k$. Our experiment shows that the model gets the best segmentation result when $k = 3$. With this architecture, O-Net can iteratively integrate attention features and predict the detail skin lesion images segmentation maps.

1) Recurrent U-Net architecture: The primary structure of the proposed network is roughly similar to parts of Recurrent U-Net, with encoders and decoders symmetrically distributed on both sides of the structure, and the previously output segmentation mask is concatenated to the input image through recurrent unit. In the encoding stage, the low dimensional feature in the input image is obtained through a rich filter, while the decoding stage, the inverse process of encoding is

carried out by up-sampling and integrate the corresponding encode layer feature information, so as to acquire the precise segmentation results of each voxel. The recurrent unit incorporated the previous segmentation mask into the input using the corresponding attention extraction fusion mechanism. Besides, in the up-sampling part, transposed convolutions are used to enhance the network's representation ability rather than bilinear interpolation. Group normalization [34] is used in partial unit to overcome the internal covariate shift (ICS) phenomenon and increase the speed of processing.

2) Attention Class Feature Module: In general, the main challenge of skin lesion images segmentation is to distinguish the low contrast voxels in the boundary regions. Figure 3 shows the challenging skin lesion images from ISIC-2017 dataset. Traditional non-learning-based approaches guide the lesion segmentation process through obtain spatial color distribution of the lesion image, which may bring bias because of the lowest tissue contrast arise from the edge region. Deep Learning based approaches such as Multi-stage U-Nets [22], CDNN [23], DermoNet [24] explore the lesion tissue features by directly combination of responses at multiple direction or scales, which not distinguish voxels from different classes. The attentional class feature network (ACFNet) solved this problem by introducing attentional class feature module into the normal convolutional neural networks. It can be seen from the characteristics of ACFNet that each voxel can combine different class centers, which are distinguish voxels from different classes explicitly. Therefore, we integrated the attentional class feature module into the proposed network.

In the attention class feature module, richer global context is explored from view of categorical through class center. The features of all voxels remain with lesion category is averaged to obtain the class center of lesion category. Therefore, the attention class feature is able to supervise the whole segmentation process from the overall presentation of each class.

In the process of the attention class feature extraction, for the input feature map $X \in \mathbb{R}^{W\times H\times C}$, in which W, H and C represent width, height and the number of channels respectively, the class center F of c category is formalized as follows:

$$F^{c}_{\text{class}} = \frac{\sum_{i=0}^{WH} \gamma\,(y_i, \text{c}) \cdot F_i}{\sum_{i=0}^{WH} \gamma\,(y_i, \text{c})} \tag{1}$$

where y_i is the actual class of voxel i and $\gamma(y_i, c)$ denotes the indicator function of binary that recognize whether the providing voxel belong to the class c.

In our approach, we use the feature map $F \in \mathbb{R}^{W\times H\times C}$ in the second layer of the encoder and coarse segmentation result $F_{coarse} \in \mathbb{R}^{W\times H\times N}$, where N denotes the number of classes, generated by encoder and decoder to calculate class center for each class. For purpose of lessen the costs of calculate the class center, for feature map F, we employ a 1×1 convolution to change its channel dimension to C'. Next, we reshape and transpose Y_{coarse} to $\mathbb{R}^{N\times WH}$ and only reshape the dimension reduction of feature map F' to $\mathbb{R}^{WH\times C'}$. Then we calculate the class center $F_{class} \in \mathbb{R}^{N\times C'}$ as follows:

$$F^{c}_{\text{class}} = \frac{\sum_{i=0}^{WH} P^{i,c}_{\text{coarse}} \cdot F'_i}{\sum_{i=0}^{WH} P^{i,c}_{\text{coarse}}} \tag{2}$$

where $P^{i,c}_{coarse}$ is the probability of voxel i belonging to class c. Both F^{c}_{class} and F'_i are in $\mathbb{R}^{1\times C'}$.

In order to monitor the update of model parameters during training, we utilize the binary cross entropy loss to measure the difference between the predicted coarse segmentation result and the corresponding ground truth, which computed on the pixel-level. The total loss is the sum of all coarse segmentation loss. The loss can be obtained from the following:

$$L_{BCE} = -\frac{1}{M}\sum_{i=1}^{M}[(1-y_i)log(1-\hat{y}_i) + y_i log\hat{y}_i] \tag{3}$$

where M denote the number of all pixels in one sample, y_i and $\hat{y}_i$ stand for the ground truth and the predict probability of pixel i, respectively.

C. COMPARE WITH RECURRENT U-NET AND ATTENTION CLASS FEATURE NETWORK (ACFNet)

Two state-of-the-art works were used to compare with our proposed method. One is the Recurrent U-Net, which three layers to englobe; The other is the attention class feature network (ACFNet). Recurrent U-Net introduce a simplified Gate Recurrent Unit (GRU) [35] dubbed Single-gated Recurrent Unit (SRU) to integrate recursions. Our proposed O-Net provides an attention module with the ability to extract attention class feature information as a recurrent unit. Compared with SRU, the attention class feature module is beneficial to distinguish the low contrast voxel at the edge of skin lesions. ACFNet focused its attention on Cityscapes [36] dataset with large training samples and achieved good segmentation results. For skin lesion images segmentation with fewer training datasets, O-Net is based on a more compact U-Net that can use attention mechanisms to achieve a coarse-to-fine segmentation effect without using a large number of parameters.

D. PERFORMANCE EVALUATION METRICS

In the experiment, we used the following metrics to evaluate our model: Accuracy (ACC), Dice coefficient (DICE), Jaccard index (JAC), Sensitivity(SEN), and Specificity(SPEC). The accuracy represents the proportion of the number of correctly predicted pixels in the picture to all pixels, which can measure the ability of the model in pixel classification. Dice and JAC are defined to measure the degree of overlap between the predicted area and the correctly labeled area, and the value is between [0, 1]. The larger the value, the closer the predicted lesion area is to the real lesion area. Because the lesion area accounts for less than normal skin in the skin lesion image, a large number of background pixels make the accuracy higher, but the actual segmentation effect is poor. Sensitivity and specificity will distinguish

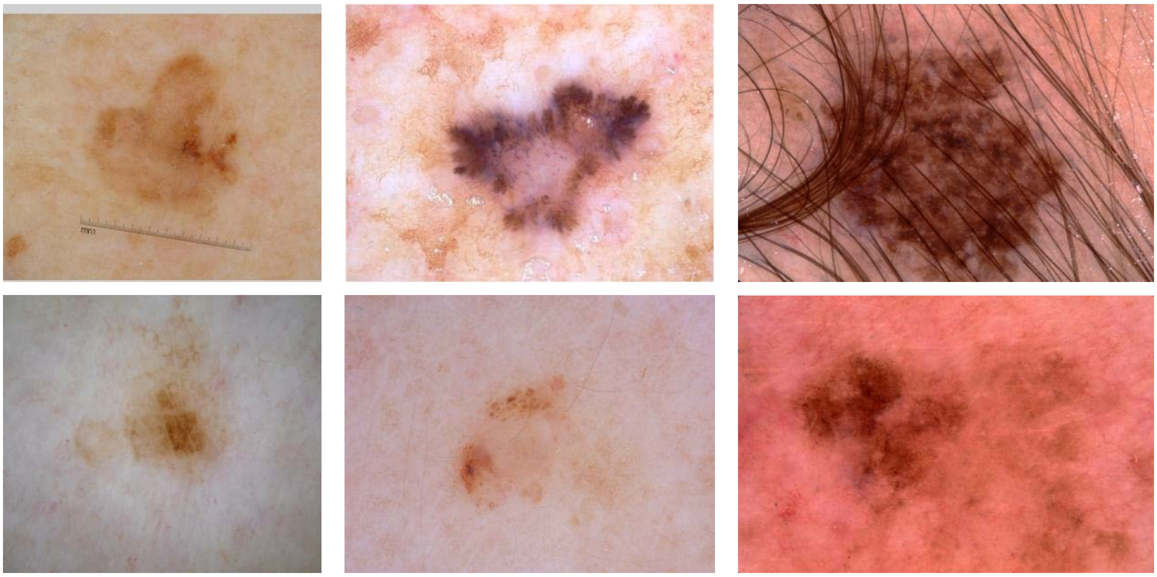

FIGURE 3. Sample images of skin lesion with low boundaries contrast from ISIC-2017 dataset, including samples with hair and marks covering the lesion area, as well as complex samples with different skin color, lesion location, lesion size, lesion shape and so on.

between normal skin and lesion areas to avoid being affected by unbalanced data.The expressions are as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$DICE = \frac{2 \times TP}{FP + FN + 2 \times TP} \tag{5}$$

$$JAC = \frac{TP}{TP + FN + FP} \tag{6}$$

$$SEN = \frac{TP}{TP + FN} \tag{7}$$

$$SPEC = \frac{TN}{TN + FP} \tag{8}$$

where TP, FP, TN, and FN represent the quantity of true positive, false positive, true negative, and false negative voxels in the binary segmentation results.

IV. EXPERIMENTAL

The proposed O-Net has recurrent unit to adaptively capture the attentional and contextual information. It enables integration of the attentional class features in the rough segmentation results to the recurrent network and obtains the information differences between the classes, so that the precise segmentation results are obtained. In this section, we systematically compared our proposed O-Net with ACFNet and Recurrent U-Net. First, the results ground on the validation set are presented and employed in the selection of hyperparameter. Then, we provide experimental results of our model and comparison model on two test sets. We also make a brief comparison between O-Net with other recently published methods, most of which are ground on the framework of convolution neural network. All experiments were carried out based on the PyTorch frameworks using an NVIDIA GeForce RTX 2080Ti GPU.

A. DATA PREPROCESSING

Due to the limited hardware resources, it is impossible to set a large training batch size when using high-resolution images during training. Moreover, it is various of the resolution of skin lesion images in ISIC dataset. Therefore, in order to enable the model to be trained in batch-wise, all images are uniformly reduced to resolution 341 × 256 pixel before being input into the network. Furthermore, to enrich the color information in lesion images, we superimposes the HSV color space on the original RGB color space so as to form a six channel image. HSV refers to defining the color space through hue, saturation and lightness. Compared with traditional RGB images, HSV images are less affected by light, and insensitive to changes in skin curvature, and can effectively improve the accuracy of skin lesion image segmentation. HSV color space is intuitive in distinguishing lesion areas from normal skin, which is beneficial to segmentation.

In addition, due to the limitation of the number of training data, we utilize several data augmentation skills to enrich the training samples and meanwhile avoid the phenomenon of overfitting. This paper mainly adopts the following data enhancement methods: (1) horizontal and vertical inversion. (2) fuzzy. (3) random rotation. (4) affine transformation. (5) random masking. (6) mesh distortion.

B. COMPARISONS WITH ACFNet AND RECURRENT U-NET

On the basis of ISIC-2017 dataset and PH2 dataset, three models of O-Net, ACFNet, Recurrent U-Net were compared. As mentioned before, the dataset was divided into three subsets: training, validation and test. We initialized the model parameter weights with random values and trained the above three models from scratch on the training set. In order to avoid overfitting and ensure a quick convergence, we utilize

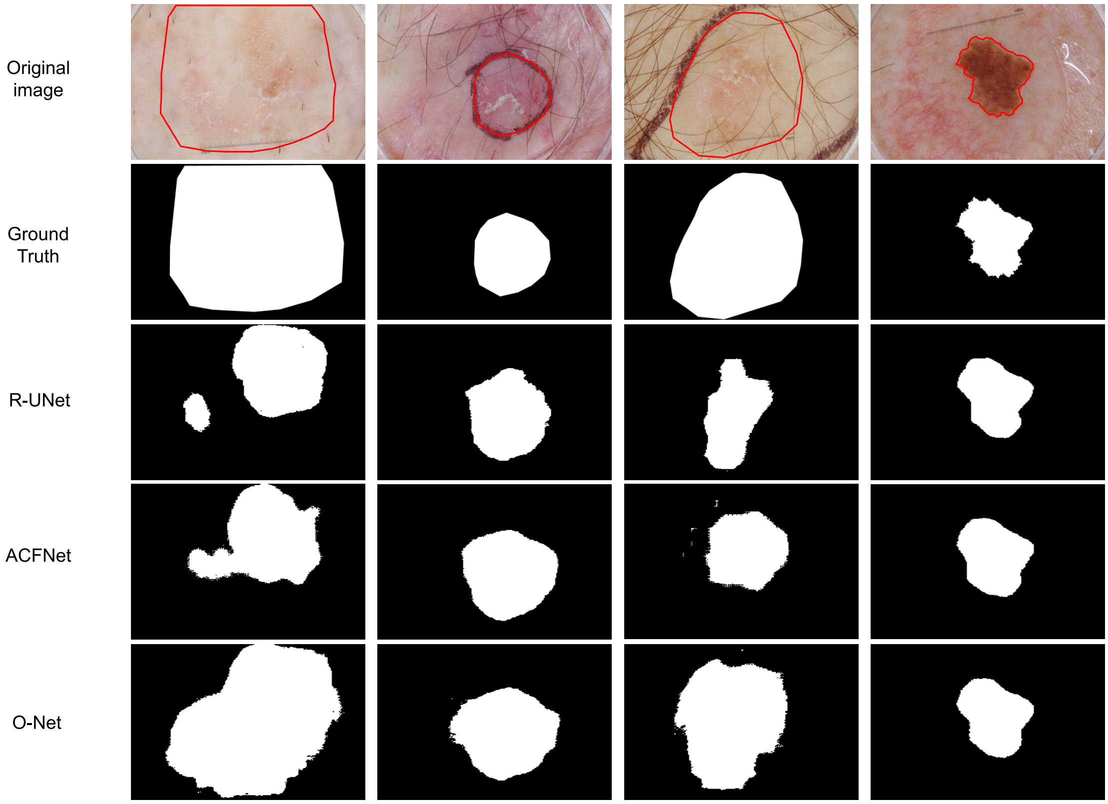

FIGURE 4. The segmentation results predicted by our model and the comparison model on ISIC-2017 dataset, each column represents a sample. The first row and the second row is the original image and the skin lesion segmentation ground truth, respectively and the remaining rows are the visual segmentation results of different models.

TABLE 2. Performance of three models trained on ISIC-2017 dataset.

Model	ACC	LOSS
R-UNet	0.9214	0.1143
ACFNet	0.9347	0.1435
O-Net(ours)	**0.9473**	**0.0927**

TABLE 3. Performance of three models trained on ISIC-2017 dataset.

Model	SEN	SPEC	ACC	DICE	JAC
R-UNet	0.8678	0.9222	0.9268	0.8314	0.7386
ACFNet	0.8513	0.9132	0.9345	0.8407	0.7497
O-Net(ours)	**0.8970**	**0.9630**	**0.9471**	**0.8704**	**0.8036**

the Adam optimizer to update parameters, the initial learning rate is set to 1e-3, momentum parameters b1=0.9, b2=0.99. Meanwhile, the learning rate decay strategy ReduceLROn-Plateau is used to dynamically adjust the learning rate during the train process, specifically, when Jaccard index does not increase after 30 iterations, the learning rate will be reduced to 1/10 of the original. The batch size is set to 8.

In the process of model training, we recorded the loss values and validation accuracy. Table 2 reflects the performance on the validation set. As shown in Table 2, ours O-Net has the highest accuracy rate and the lowest loss value of 0.9473 and 0.0927, respectively on ISIC-2017 dataset.

We used the test set for the further evaluation of the model. Table 3, Table 4 show the comparison of SEN, SPEC, ACC, DICE and JAC on the three model. Compared to the other methods, O-Net obtains the best performance on all metrics as shown in the tables. It's worth noting that the O-Net achieves the highest average Dice coefficient among the several models. The average DICE for R-UNet, ACFNet and O-Net is 0.8314/0.8407/0.8704 on ISIC-2017 dataset, 0.9057/0.9176/0.9212 on PH2 dataset, respectively.

TABLE 4. Performance of three models tested on PH2 dataset.

Model	SEN	SPEC	ACC	DICE	JAC
R-UNet	0.8425	0.9365	0.9412	0.9057	0.8371
ACFNet	0.8310	0.9513	0.9205	0.9176	0.8567
O-Net(ours)	**0.8923**	**0.9675**	**0.9514**	**0.9212**	**0.8615**

C. SKIN LESION IMAGE SEGMENTATION RESULTS

The results of skin lesion image segmentation were shown in Figure 4 and Figure 5. The figure shows lesion segmentation results produced by O-Net were more refined at the lesion edge. The proposed O-Net can distinguish low-contrast parts which may be confused in ACFNet and Recurrent U-Net, thus it allows more details to be preserved.

Figure 6 show the difference of segmentation results generated by each model in detail, which provides a typical view of skin edge regions where the voxel intensity distribution of skin and lesion is highly overlapping. Due to the low tissue contrast at the lesion edge, it is difficult for segmentation algorithms to accurately process such tissue. Due to the limitation of model architecture, ACFNet and Recurrent U-Net

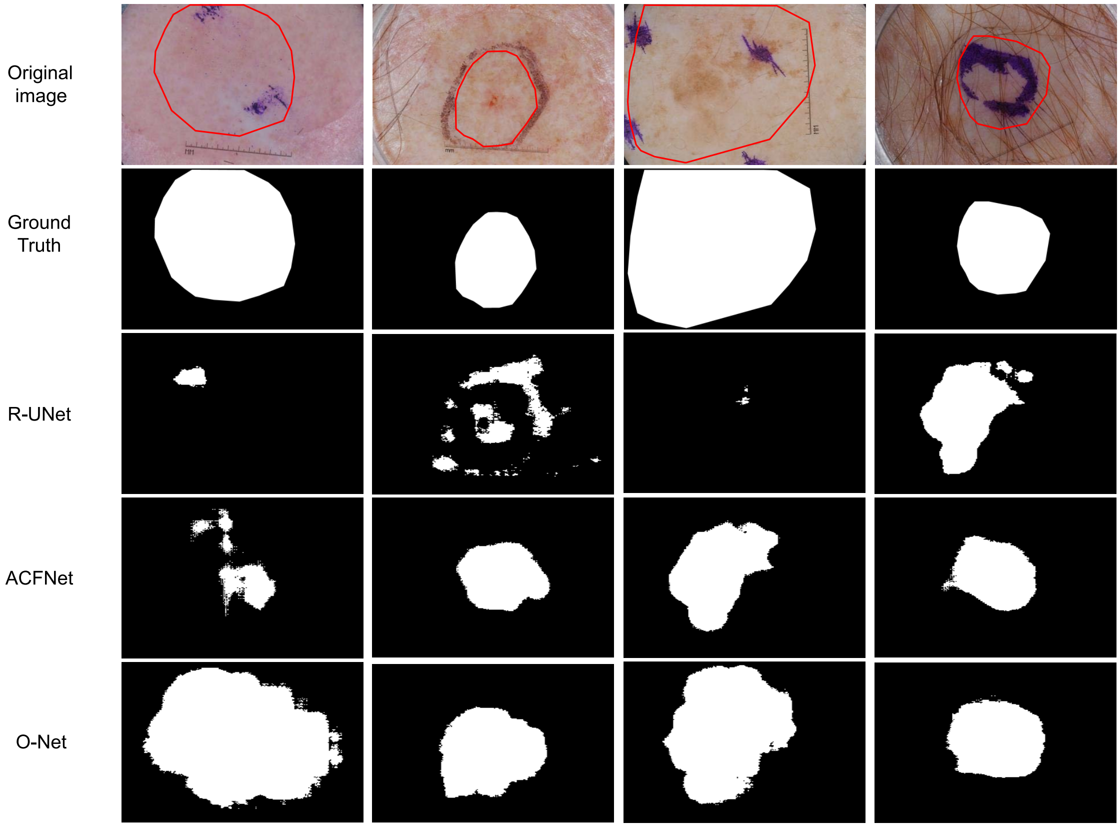

FIGURE 5. Representative segmentation results predicted by our model and the comparison model. Each column represents a sample. The first row and the second row are the original image and the skin lesion segmentation ground truth, respectively and the remaining rows are the visual segmentation results of different models.

extracted incomplete information in the edge regions of the skin lesion. It is noteworthy that ACFNet extracted more detailed lesion than Recurrent U-Net in edge regions, which presented its capability to distinguish different class of low signal differentiation. With the aid of attention class feature module, the O-Net successfully segmented the overlapping voxels at the lesion edge. In low intensity distribution of voxels regions, Recurrent U-Net shows limitations in dealing with the details. However, ACFNet had identified them correctly in some places. Therefore, the O-Net obtained desired segmentation results in overlapping and low contrast tissue.

The experimental results present that the novel network architecture characteristics of O-Net have better performance compared with among the other three models, especially, when dealing with complex and low-contrast lesion voxels.

D. COMPARISON AGAINST EXISTING METHODS

Our approach is also compared with several state-of-the-art approaches. Among them, some are traditional convolution neural network approaches while the others are adversarial learning approaches. We generalize the performance on ISIC-2017 and PH2 dataset in Table 5 and Table 6. The data show that O-Net architecture almost achieves the highest metrics among all methods on ISIC-2017 dataset. It produces the highest ACC with 0.9471, which represents that the O-Net achieves best performance comparing with convolutional neural network based methods and adversarial learning methods. Furthermore, the confusion matrices on ISIC-2017 and PH2 datasets are shown in Figure 7. Although the O-Net does not perform much better than other approaches on PH2 dataset, the results it produces are still competitive and demonstrate its strong generalization ability.

TABLE 5. Comparison with state-of-the-art networks on ISIC-2017 dataset.

Method	SEN	SPEC	ACC	DICE	JAC
Al-Masni et al. [38]	0.8540	0.9669	0.9403	0.8708	0.7711
Bi et al. [37]	0.8620	0.9671	0.9408	0.8566	0.7773
Lei et al. [39]	0.8350	0.9760	0.9350	0.8590	0.7710
Mirikharaji et al. [40]	0.8550	0.9730	0.9380	0.8570	0.7730
Sarker et al. [41]	0.8160	**0.9830**	0.9360	**0.8780**	0.7820
ASCU-NET [28]	0.8250	0.9650	0.9260	0.8300	0.7420
Zhang et al. [31]	0.8524	0.9858	0.9419	0.8723	0.7831
O-Net(ours)	**0.8970**	0.9630	**0.9471**	0.8704	**0.8036**

TABLE 6. Comparison with state-of-the-art networks on PH2 dataset.

Method	SEN	SPEC	ACC	DICE	JAC
Al-Masni et al. [38]	0.9372	0.9565	0.9508	0.9177	0.8479
Bi et al. [42]	0.9489	0.9389	0.9424	0.9066	0.8399
Peng et al. [43]	0.8700	**0.9700**	0.9300	0.9000	0.8500
ASCU-NET [28]	**0.9600**	0.9370	0.9430	0.9090	0.8420
O-Net(ours)	0.8923	0.9675	**0.9514**	**0.9212**	**0.8615**

Although the dice coefficient of O-Net performs no better than Sarker *et al.*'s method [41] on ISIC-2017 dataset, O-Net is ahead in both accuracy and Jaccard index, which indicate

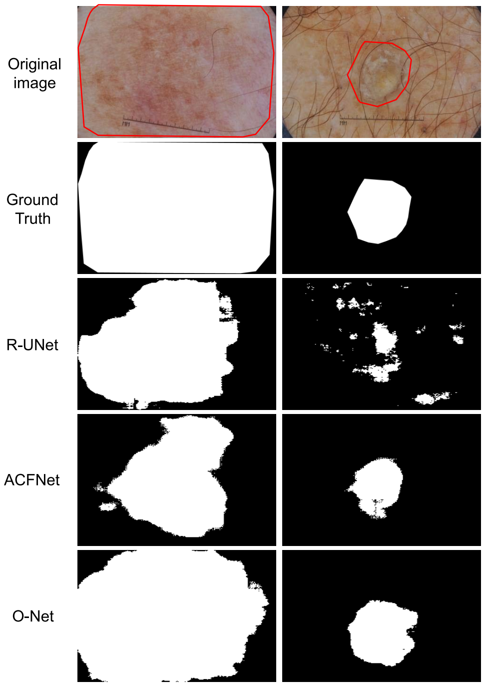

FIGURE 6. The detail Segmentation results predicted by our model and the comparison model. Samples shown here contains lesions in extreme cases. Similarly, the first row shows the original images, and the second row is the skin lesion segmentation ground truth, and the remaining rows are the visual segmentation results of different models.

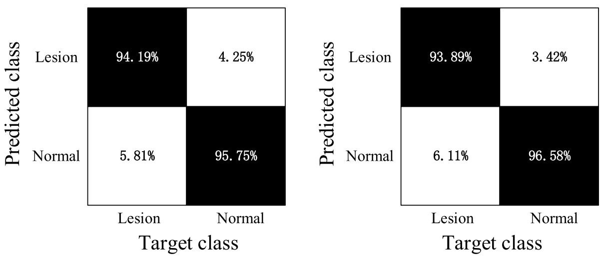

FIGURE 7. The confusion matrix on ISIC-2017(left) and PH2(right) datasets.

our method improves the accuracy while maintaining the integrity of the lesion regions. Similarly, when our model is extended to PH2 dataset, despite the fact that Bi *et al.*'s model [37] achieves higher accuracy, its ability to correctly distinguish the lesion region from normal skin is inadequate, so it is lower than O-Net on dice coefficient and Jaccard index. The reason may be the result of O-Net's comprehensive consideration of attentional class feature and lesion feature.

E. ANALYSIS OF THE INFLUENCE OF THE BACKBONE

The proposed O-Net is essentially composed of an encoder-decoder structure and a recurrent attention mechanism module AFFM, which captures context information and achieves feature fusion with iterative refinement. Intuitively, any network with U-Net structure can be used as a backbone to replace the original U-Net model used in the experiments. Here, we choose two different common used backbone networks: VGG16 and ResNet50. The experiments are conducted on the ISIC-2017 dataset, and the results are shown in Table 7. It can be seen that, AFFM achieves competitive performance with different backbones, which shows the great generalization ability. Specifically, AFFM shows better segmentation performance on ResNet50 compared with U-Net and VGG16. It can be explained that, utilizing the backbone with stronger feature extraction ability is more conducive to AFFM to further achieve feature fusion and recurrent optimization.

TABLE 7. Comparison with different encoder networks on ISIC-2017 dataset.

Method	SEN	SPEC	ACC	DICE	JAC
U-Net+AFFM	0.8970	0.9630	0.9471	0.8704	0.8036
VGG16+AFFM	0.8825	0.9469	0.9403	0.8523	0.7811
ResNet50+AFFM	0.8942	0.9629	0.9472	0.8716	0.8150

F. ANALYSIS OF INCREASING THE RECURSIVE ITERATION

In this study, the relationship between recursive iterations K and model performance is studied in detail. For this purpose, we use different iterations to train our model, i.e., $K = 1, 2, \ldots, 6$. Figure 8 describes how the performance of the network changes with the number of iterations. The performance has improved significantly when the number of iterations increases from 1 to 3. Meanwhile, we visualize the segmentation performance in Figure 9. It can be seen that, as the iteration K increasing, the segmented lesion area gradually approaching to the Ground Truth, which can be proved that the proposed recurrent attention mechanism capturing more detailed attention feature information through input coarse-to-fine segmentation result. However, the performance has been no further significant improvement, when the number of iterations was further increased. One possible reason is when the number of iterations more than 3, the limited the basic network structure and the integration of attention class features, the rough segmentation results cannot be further improved. Therefore, in our experiments, for better balance the efficiency and effectiveness, we set 3 as the value of iterations K. Improving integration of attention class features will be the focus of our follow-up research.

V. DISCUSSION

Attention mechanism, which inspires by the way of humans acquire scene information, have been generally used for semantic segmentation tasks and achieved remarkable performance. O-Net is extended by introducing attention class feature module as recurrent fusion unit on the idea of the Recurrent U-Net. With the circular O-shape architecture, O-Net is designed to generate coarse segmentation result by standard U- Net and integrate the attention class feature information by recurrent unit through extracted in the coarse segmentation into the encoder. Furthermore, with the attention class feature module, O-Net is able to capture the information differences between classes by combine different class centers. By using the coarse skin lesion segmentation result

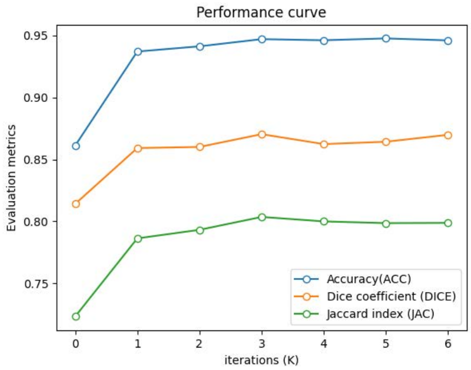

FIGURE 8. The performance of the model with different iterations.

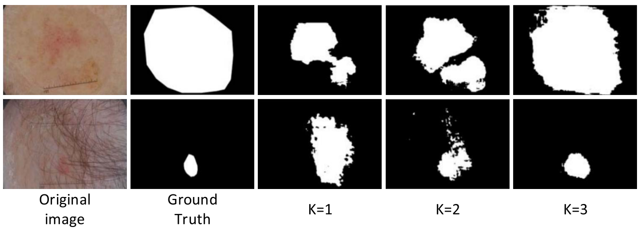

FIGURE 9. The segmentation performance of different iterations.

of each iteration as attention feature map of the next iteration to capture attentional class feature, the different pixels focus on different classes.

Begin, we have trained Recurrent U-Net and ACFNet from scratch to compare the performance with our proposed approach. It is also the first time O-Net has been used to handle the skin lesion segmentation. Besides, O-Net is compared with several deep convolutional network based approaches and some other non-learning-based approaches.

Although O-Net complete the skin lesion image segmentation task to some extent, there are still aspects worthy of exploration. On the one hand, O-Net only segments apart the normal skin and lesion area, but lacks of further division on lesion types, which only provides limited information for subsequent process. Therefore, it is necessary to explore how to further divide the specific area in combination with the clinical diagnostic criteria for the diagnosis of pigmented skin lesions, so as to provide more valuable information for doctors' diagnosis and subsequent analysis. On the other hand, the training of O-Net needs to manually set the number of iterations and other hyperparameters, which is time-consuming and only achieve suboptimal in general. In subsequent research, the neural network automatic search (NAS) technology can be utilized to automatically explore the optimal parameters for the skin lesion image segmentation task.

Moreover, O-Net utilizes the basic encoder and decoder structure to introduce long-distance class-level attention information through recurrent iteration, which realizes the accurate segmentation of skin lesion area and ensures the compactness and lightweight of the model. Combined with the characteristics of relatively simple and small datasets, the compact and lightweight framework is universal in the field of medical image processing. Therefore, we aim to explore the application of the network proposed in this paper in other medical image fields.

VI. CONCLUSION

In the study, we build a recurrent attentional convolutional network, named O-Net to deal with the skin lesion images segmentation task. O-Net is equipped with the attention feature fusion module in the recurrent unit, which iteratively compromise the attention mechanism and capture sufficient context information to refine the skin lesion segmentation results. In experiments, the model is trained and tested on two datasets: ISIC-2017 dataset and PH2 dataset. Results show that with the help of attention class feature module, more detailed lesion regions are extracted, and achieved competitive performance in segmenting the skin lesion images.

REFERENCES

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, Jan. 2017.

[2] (2020). *American Cancer Society, Key Statistics for Melanoma Skin Cancer*. [Online]. Available: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html

[3] Z. Ge, S. Demyanov, R. Chakravorty, A. Bowling, and R. Garnavi, "Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2017, pp. 250–258.

[4] L. Smith and S. MacNeil, "State of the art in non-invasive imaging of cutaneous melanoma," *Skin Res. Technol.*, vol. 17, no. 3, pp. 257–269, Aug. 2011.

[5] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Comput. Med. Imag. Graph.*, vol. 31, pp. 198–211, Jun. 2007.

[6] H. Fan, F. Xie, Y. Li, Z. Jiang, and J. Liu, "Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold," *Comput. Biol. Med.*, vol. 85, pp. 75–85, Jun. 2017.

[7] A. Jalalian, S. Mashohor, R. Mahmud, B. Karasfi, M. I. B. Saripan, and A. R. B. Ramli, "Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection," *EXCLI J.*, vol. 16, p. 113, Feb. 2017.

[8] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artif. Intell. Med.*, vol. 56, no. 2, pp. 69–90, 2012.

[9] N. K. Mishra and M. E. Celebi, "An overview of melanoma detection in dermoscopy images using image processing and machine learning," 2016, *arXiv:1601.07843*.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] F. Zhang, Y. Chen, Z. Li, Z. Hong, J. Liu, F. Ma, J. Han, and E. Ding, "ACFNet: Attentional class feature network for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6798–6807.

[12] W. Wang, K. Yu, J. Hugonot, P. Fua, and M. Salzmann, "Recurrent U-Net for resource-constrained segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2142–2151.

[13] P. G. Cavalcanti and J. Scharcanski, "Automated prescreening of pigmented skin lesions using standard cameras," *Comput. Med. Imag. Graph.*, vol. 35, no. 6, pp. 481–491, Sep. 2011.

[14] S. Pathan, K. G. Prabhu, and P. C. Siddalingaswamy, "Hair detection and lesion segmentation in dermoscopic images using domain knowledge," *Med. Biol. Eng. Comput.*, vol. 56, no. 11, pp. 2051–2065, Nov. 2018.

[15] E. Goceri and M. Gunay, "Automated detection of facial disorders (ADFD): A novel approach based-on digital photographs," *Comput. Methods Biomech. Biomed. Eng., Imag. Visualizat.*, vol. 6, no. 6, pp. 607–617, 2018.

[16] H. Zhou, X. Li, G. Schaefer, M. E. Celebi, and P. Miller, "Mean shift based gradient vector flow for image segmentation," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1004–1016, 2013.
[17] J. Glaister, A. Wong, and D. A. Clausi, "Segmentation of skin lesions from digital images using joint statistical texture distinctiveness," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1220–1230, Apr. 2014.
[18] L. Bi, J. Kim, E. Ahn, D. Feng, and M. Fulham, "Automated skin lesion segmentation via image-wise supervised learning and multi-scale superpixel based cellular automata," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1059–1062.
[19] M. Nasir, M. A. Khan, M. Sharif, I. U. Lali, T. Saba, and T. Iqbal, "An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach," *Microsc. Res. Technique*, vol. 81, no. 6, pp. 528–543, Jun. 2018.
[20] T. Y. Satheesha, D. Satyanarayana, M. N. G. Prasad, and K. D. Dhruve, "Melanoma is skin deep: A 3D reconstruction technique for computerized dermoscopic skin lesion classification," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–17, 2017.
[21] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, Dec. 2016.
[22] Y. Tang, F. Yang, S. Yuan, and C. Zhan, "A multi-stage framework with context information fusion structure for skin lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1407–1410.
[23] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1876–1886, Sep. 2017.
[24] S. Baghersalimi, B. Bozorgtabar, P. Schmid-Saugeon, H. K. Ekenel, and J.-P. Thiran, "DermoNet: Densely linked convolutional neural network for efficient skin lesion segmentation," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, pp. 1–10, Dec. 2019.
[25] L. Bi, D. Feng, M. Fulham, and J. Kim, "Improving skin lesion segmentation via stacked adversarial learning," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1100–1103.
[26] Y. Xie, J. Zhang, Y. Xia, and C. Shen, "A mutual bootstrapping model for automated skin lesion segmentation and classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2482–2493, Dec. 2020.
[27] O. O. Abayomi-Alli, R. Damaševicius, S. Misra, R. Maskeliunas, and A. Abayomi-Alli, "Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 29, pp. 2600–2614, Oct. 2021.
[28] X. Tong, J. Wei, B. Sun, S. Su, Z. Zuo, and P. Wu, "ASCU-Net: Attention gate, spatial and channel attention U-Net for skin lesion segmentation," *Diagnostics*, vol. 11, no. 3, p. 501, Mar. 2021.
[29] M. D. Alahmadi, "Multiscale attention U-Net for skin lesion segmentation," *IEEE Access*, vol. 10, pp. 59145–59154, 2022.
[30] M. Nawaz, T. Nazir, M. Masood, F. Ali, M. A. Khan, U. Tariq, N. Sahar, and R. Damaševičius, "Melanoma segmentation: A framework of improved DenseNet77 and UNET convolutional neural network," *Int. J. Imag. Syst. Technol.*, vol. 32, no. 3, pp. 30–44, May 2022.
[31] G. Zhang and S. Wang, "Dense and shuffle attention U-Net for automatic skin lesion segmentation," *Int. J. Imag. Syst. Technol.*, vol. 32, no. 4, pp. 1–14, Jun. 2022.
[32] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.
[33] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH^2—A dermoscopic image database for research and benchmarking," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 5437–5440.
[34] Y. Wu and K. He, "Group normalization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.
[35] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, *arXiv:1409.1259*.
[36] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.
[37] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," *Pattern Recognit.*, vol. 85, pp. 78–89, Jan. 2019.
[38] M. A. Al-masni, M. A. Al-Antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Comput. Methods Programs Biomed.*, vol. 162, pp. 221–231, Aug. 2018.
[39] B. Lei, Z. Xia, F. Jiang, X. Jiang, Z. Ge, Y. Xu, J. Qin, S. Chen, T. Wang, and S. Wang, "Skin lesion segmentation via generative adversarial networks with dual discriminators," *Med. Image Anal.*, vol. 64, Aug. 2020, Art. no. 101716.
[40] Z. Mirikharaji and G. Hamarneh, "Star shape prior in fully convolutional networks for skin lesion segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2018, pp. 737–745.
[41] M. M. K. Sarker, "Slsdeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2018, pp. 21–29.
[42] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic image segmentation via multistage fully convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2065–2074, Sep. 2017.
[43] Y. Peng, N. Wang, Y. Wang, and M. Wang, "Segmentation of dermoscopy image using adversarial networks," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 10965–10981, Apr. 2019.

PENG CHEN received the Ph.D. degree in medicine from Jilin University. She is currently an Associate Professor with the Department of Pediatrics, The Second Hospital of Jilin University.

SA HUANG received the Ph.D. degree in medicine from Jilin University, China. He is currently working as an Associate Professor with the Department of Radiology, The Second Hospital of Jilin University, China. His research interests include computational intelligence, imaging diagnosis, and molecular imaging.

QING YUE received the M.S. degree from the College of Computer Science and Technology, Jilin University, in 2018. She is currently a Teaching Assistant with the College of Mathematics and Computer Science, Jilin Normal University. Her current research interests include computer vision, deep learning, and data mining.